

<2024 전기 졸업과제 중간 보고서>



주제	개인의 특성 기반 패션 추천 시스템
팀명	퍼스널릭스
팀 번호	24
팀원	이성훈, 김태훈

목 차

1. 과제 배경 및 목표
 - 1.1. 과제 배경
 - 1.2. 과제 목표
 - 1.3. 관련 개념
 - 1.3.1. 퍼스널 컬러
 - 1.3.2. 얼굴형
 - 1.3.3. 체형
2. 요구조건 및 제약사항 대응전략
 - 2.1. 개인정보 보호 및 사용자의 진입장벽 문제
 - 2.2. 추천 품질 피드백
3. 설계 상세화 및 변경 내역
 - 3.1. 설계 상세화
 - 3.1.1. 사용자 입력 및 데이터 전송
 - 3.1.2. 데이터 전처리 및 분석
 - 3.1.3. 패션 추천
 - 3.1.4. 추천 결과 반환 및 사용자 피드백
 - 3.1.5. 전체 구상도
 - 3.2. 변경 내역
 - 3.2.1. 추천 알고리즘 보완
 - 3.2.2. UX 및 데이터 흐름 개선
4. 갱신된 과제 추진 계획
 - 4.1. 데이터 처리 방식
 - 4.2. 추천 알고리즘 보완
 - 4.3. UX 및 데이터 흐름 개선
 - 4.4. 개발 일정
5. 구성원별 진척도

6. 과제 수행 내용 및 중간 결과

6.1. 퍼스널 컬러 측정

6.1.1. 데이터 전처리

6.1.2. 피부 영역 RGB 값 추출

6.1.3. 색상 공간 변환 및 퍼스널 컬러 분류

6.2. 얼굴형 측정

6.2.1. 데이터 전처리

6.2.2. 모델 구성

6.2.3. 모델 학습

6.2.3.1. EfficientNetB4를 제외한 나머지 학습

6.2.3.2. 전체 학습

6.3. 체형 측정

6.3.1. KeyPoint 추출

6.3.2. 허리 위치 추측

6.3.3. 가슴 위치 추측

6.3.4. 배경 제거 및 길이 측정

6.3.5. 체형 측정

6.4. 패션 추천

6.4.1. RandomForest를 활용한 추천시스템

6.4.1.1. 데이터 전처리

6.4.1.2. Random Forest를 활용한 학습

6.4.1.3. 추천 품질에 대한 피드백

6.4.2. 콘텐츠 기반 협업필터링 추천시스템

6.4.2.1. 데이터 전처리

6.4.2.2. Tf-Idf Vectorizing

6.4.2.3. 코사인 유사도 구하기

6.4.2.4. 사용자의 신체적 특징에 어울리는 옷 특성 연결

1. 과제 배경 및 목표

1.1 과제 배경

현대 사회에서 패션은 개인의 개성을 표현하는 중요한 수단입니다. 그러나 모든 사람에게 적합한 패션을 찾기란 쉽지 않습니다. 특히, 사용자의 퍼스널 컬러, 얼굴형, 체형 등의 신체적 특성을 고려한 맞춤형 패션을 제안하는 시스템은 드뭅니다. 또한, 최근 개인 맞춤형 서비스의 수요가 증가하고 있습니다. 이는 사람들이 자신만의 개성을 표현하고자 하는 욕구가 커짐에 따라, 개개인의 특성을 반영한 맞춤형 서비스를 원하는 경향이 강해지고 있음을 반영합니다. 이러한 필요성을 바탕으로 우리는 '사용자의 신체적 특성을 고려한 패션 추천 시스템'을 개발하고자 합니다.



그림 1. 개인화 서비스에 대한 소비자 태도 조사¹⁾

1.2 과제 목표

본 졸업 과제의 목표는 다음과 같습니다:

- 사용자의 신체적 특성을 토대로 맞춤형 패션을 추천하는 알고리즘 개발.
- 상의, 하의, 외투 등으로 구성된 하나의 완성된 패션을 추천하는 시스템 구축.
- 사용자 친화적인 인터페이스와 높은 정확도를 갖춘 패션 추천 시스템 제공.

1) 2023년 소비 트렌드 시리즈 - 03 개인맞춤형 서비스.

https://www.mezzomedia.co.kr/data/insight_m_file/insight_m_file_1605.pdf

1.3 관련 개념

1.3.1 퍼스널 컬러

퍼스널 컬러는 개인이 타고날 대부터 가지고 있는 신체(피부, 머리카락, 눈동자) 색을 의미한다. 퍼스널 컬러는 크게 봄, 여름, 가을, 겨울의 4가지 유형으로 구분할 수 있다. 봄, 가을 유형은 워밍(warm)톤으로 분류되며, 여름, 겨울 유형은 쿨(cool)톤으로 분류된다.

1.3.2 얼굴형

얼굴형이란 얼굴의 외형적인 모양을 말하며, 각 사람마다 얼굴의 뼈 구조와 피부의 두께, 얼굴의 근육 등에 따라 다르게 나타난다. 얼굴형은 하트형, 긴형, 타원형, 둥근형, 사각형으로 분류된다.

1.3.3 체형

체형은 신체의 전체적인 형태와 구조를 말하며, 뼈의 구조, 근육의 분포, 지방의 분포 등으로 결정된다. 남자 체형은 삼각형, 역삼각형, 타원형, 사각형, 사다리꼴 형으로, 여자 체형은 역삼각형, 직사각형, 둥근형, 삼각형, 모래시계형으로 분류될 수 있다.

2. 요구조건 및 제약사항 대응전략

2.1 개인정보 보호 및 사용자의 진입장벽 문제

사용자가 체형, 얼굴형, 퍼스널 컬러 등의 민감한 개인정보를 제공해야 하는 시스템의 경우, 사용자가 참여할 진입장벽이 높아질 수 있습니다. 특히, 이런 정보 제공에 대한 불안감으로 참여를 꺼릴 가능성이 있습니다.

- 데이터 익명화: 사용자 데이터를 익명화하여 수집하고 처리함으로써 개인정보 유출 위험을 최소화합니다. 사용자의 신체적 특성을 포함한 데이터를 익명화하고, 그 데이터를 기반으로만 추천을 생성하여, 사용자의 개인정보가 노출되지 않도록 보장합니다.
- 선택적 정보 제공: 사용자가 반드시 모든 민감한 정보를 제공하지 않더라도 추천을 받을 수 있도록 시스템을 설계합니다. 초기 설정 시 기본적인 정보를 제공한 후, 사용자가 점진적으로 더 많은 정보를 제공할 때 더욱 개인화된 추천을 받을 수 있도록 하여 참여 유도를 극대화합니다.
- 데이터 투명성: 사용자에게 정보 제공의 목적과 데이터 사용 방법을 명확히 알리고, 언제든지 데이터를 수정하거나 삭제할 수 있는 권한을 제공합니다. 또한, 제공된 정보가 얼마나 정확한 추천으로 연결되는지 사용자에게 피드백을 제공하여 데이터 제공의 가치를 직관적으로 느낄 수 있게 합니다.

2.2 추천 품질 피드백

사용자의 피드백을 반영하여 추천 품질을 지속적으로 개선할 수 있는 체계가 필요합니다.

- 피드백 루프 설계: 사용자로부터 제공된 피드백(예: 별점 평가 등)을 추천 시스템에 반영하는 루프를 설계합니다. 사용자가 평가한 옷에 대한 피드백 데이터를 기존의 학습 데이터에 추가하고, 이를 활용하여 모델을 재훈련시킵니다. 이를 통해 사용자 맞춤형 추천 품질이 지속적으로 개선됩니다.
- 가중치 조정: 피드백 반영 시, 사용자의 별점 평가가 일정 비율로 반영되도록 가중치를 설정하여 추천 결과가 적절히 조정되도록 합니다. 예를 들어, 사용자 피드백이 반영된 결과와 기존 알고리즘 예측 결과를 70:30 비율로 조합하는 방식 등을 적용하여 최종 추천을 도출합니다.

3. 설계 상세화 및 변경 내역

서비스의 전체적인 흐름은 사용자가 사진 또는 신체적 정보를 업로드하면 이를 백엔드로 전달하여 이미지 분석 및 신체적 특징을 추출하고, 이를 바탕으로 패션을 추천한 뒤 결과를 프론트엔드에 전달하는 방식으로 진행됩니다.

3.1 설계 상세화

3.1.1 사용자 입력 및 데이터 전송

사용자는 사진(얼굴, 전신) 또는 신체적 정보를 입력할 수 있습니다. 이 정보는 안드로이드 앱에서 사용자의 입력 형태에 따라 사진 파일로 변환되거나 텍스트 정보로 입력됩니다. 이후, 안드로이드 프론트엔드는 사용자의 데이터를 HTTP 요청으로 백엔드에 전달합니다. 이 과정에서 데이터는 JSON 형식 또는 이미지 파일 형식으로 전송됩니다.

3.1.2 데이터 전처리 및 분석

전달된 이미지가 있을 경우, 백엔드에서는 신체적 특징 분석 모델을 사용하여 퍼스널 컬러, 얼굴형, 체형의 특징을 추출합니다. 이후, 추출된 특징은 DB에 저장합니다.

3.1.3 패션 추천

분석 또는 입력된 신체적 특성을 바탕으로, 학습된 모델을 사용하여 적합한 패션 스타일을 예측합니다. 모델은 다양한 사용자 피드백을 바탕으로 성능을 지속적으로 개선하며, 추천 품질을 향상시킵니다.

3.1.4 추천 결과 반환 및 사용자 피드백

추천된 패션 정보는 백엔드에서 다시 프론트엔드로 전달되며, 사용자는 해당 추천 카테고리나 관련 의상 이미지를 볼 수 있습니다. 사용자가 추천받은 의상에 대한 피드백을 제공하면, 이를 백엔드에서 저장하고 머신러닝 모델의 재학습에 반영하여 향후 추천 품질을 개선합니다.

3.1.5 전체 구상도

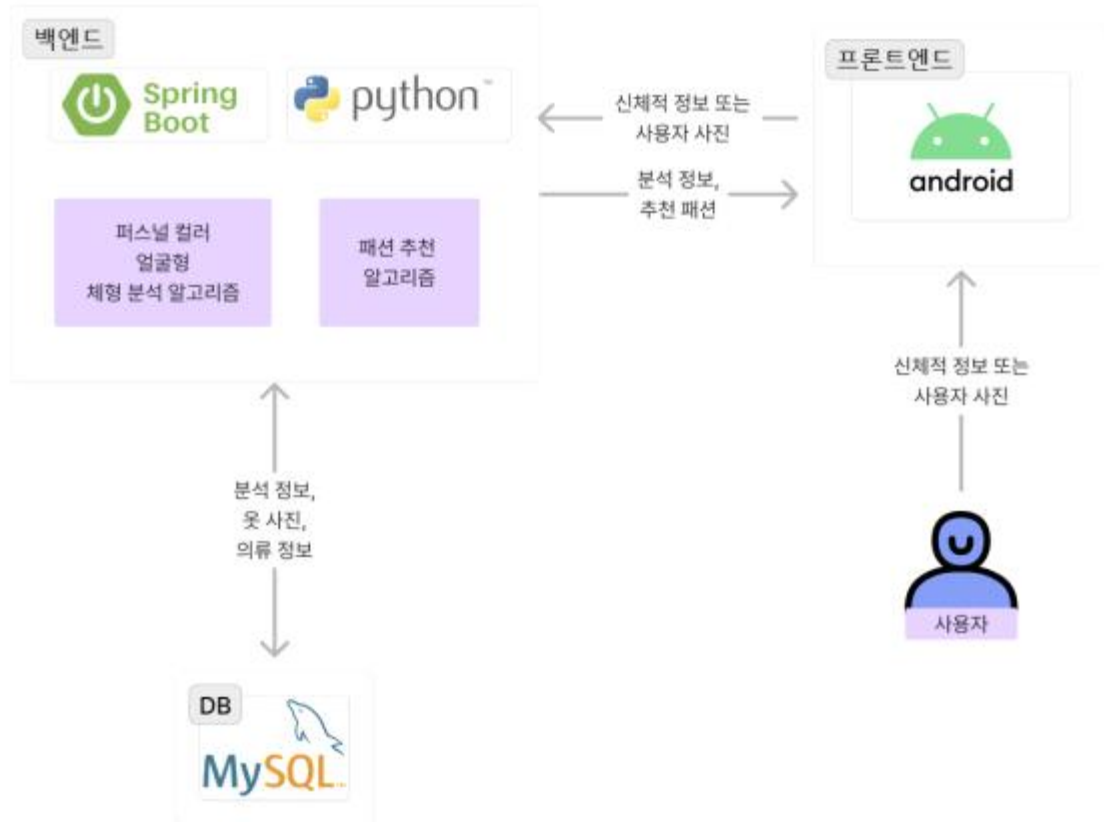


그림2. 전체 구상도

3.2 변경 내역

3.2.1 추천 알고리즘 보완

추천 품질을 높이기 위해 사용자 피드백을 반영하는 피드백 루프를 추가하여, 사용자의 평가 데이터를 지속적으로 모델 학습에 반영하도록 변경하였습니다. 이는 사용자 경험을 지속적으로 개선하고, 맞춤형 패션 추천의 정확도를 향상시키는 방향으로 설계되었습니다.

3.2.2 UX 및 데이터 흐름 개선

사용자가 신체적 정보를 직접 입력하는 방식 외에도, 사진을 통한 자동 분석 기능을 추가하여 사용자의 편의성을 높였습니다. 이에 따라 사용자 흐름과 데이터 전송 방식도 개선되었습니다. 사용자는 사진을 업로드하면 자동으로 분석이 이루어지고, 결과를 시각적으로 확인할 수 있게 되었습니다.

4. 갱신된 과제 추진 계획

본 프로젝트는 사용자 신체적 특성을 기반으로 맞춤형 패션 추천 시스템을 구축하는 것을 목표로 하고 있습니다. 프로젝트의 원활한 진행을 위해 초기 설계에서 발생한 제약사항을 해결하고, 변경된 요구사항에 맞춘 개발 전략을 반영한 갱신된 추진 계획은 다음과 같습니다.

4.1 데이터 처리 방식

추천 결과는 카테고리 정보와 관련 의상 이미지로 제공됩니다. 이미지뿐만 아니라 카테고리 정보도 제공함으로써 사용자가 원하는 의상이 아닌 경우, 외부 쇼핑몰을 통해 다른 의상을 탐색할 수 있는 기회를 제공하고자 합니다.

4.2 추천 알고리즘 보완

사용자 피드백을 반영하는 피드백 루프를 추가할 예정입니다. 사용자가 추천받은 의상에 대해 별점 평가를 하면, 이 데이터를 모델 학습에 반영하여 추천 정확도를 개선하는 구조입니다. 이를 통해 시간이 지남에 따라 추천 품질이 지속적으로 개선됩니다.

4.3 UX 및 데이터 흐름 개선

기존에는 사용자가 신체적 정보를 직접 입력해야 했으나, 사용자가 사진을 업로드하면 자동으로 신체적 특성을 분석하는 기능을 추가하였습니다. 머신러닝 기반 이미지 분석 모델을 통해 사용자의 얼굴형, 체형, 퍼스널 컬러를 자동으로 추출하고, 이를 기반으로 패션 추천이 이루어집니다. 이 기능을 통해 사용자 편의성이 대폭 개선됩니다.

4.4 개발 일정

6월					7월					8월					9월					10월		
1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3
추천 알고리즘 스터디																						
				데이터 수집 및 전처리																		
					신체적 특성 분석 알고리즘 개발																	

5. 구성원별 진척도

이성훈:

- 데이터 수집 및 전처리 완료
- 신체적 특징 분석 모델 개발 완료
- 백엔드 및 프론트엔드 제작 중
- 모델 수정 및 보완 중

김태훈:

- 데이터 수집 및 전처리 완료
- 신체적 특징 분석 모델 개발 완료
- 추천 모델 개발 완료
- 모델 수정 및 보완 중

6. 과제 수행 내용 및 중간 결과

6.1 퍼스널 컬러 측정

6.1.1 데이터 전처리

사용된 데이터셋은 AIHub의 한국인 안면 이미지²⁾로, 약 3만장의 이미지로 구성됩니다. 이 중에서 정면사진만을 사용하여, 약 5000개의 이미지를 사용하였습니다. 각 이미지의 얼굴의 피부 영역만을 추출하여 정확한 퍼스널 컬러 측정을 진행했습니다.

얼굴에서 피부 영역만을 추출하기 위해 딥러닝 기반의 얼굴 탐지 및 세분화 모델을 적용하였습니다. facer 라이브러리의 retinaface/mobilenet 모델을 사용하여 얼굴을 감지한 후, farl/lapa/448 모델을 활용하여 얼굴을 세분화했습니다. 이 과정에서 피부 영역을 정확히 구분하기 위해 얼굴의 각 부분(눈, 코, 입, 이마, 피부 등)을 확률 맵으로 분할하고, 피부 부분만을 선택하였습니다.



그림 3. 원본 사진(좌), 추출된 마스크(우)

6.1.2 피부 영역 RGB 값 추출

세분화된 확률 맵에서 피부 부분에 해당하는 이진 마스크를 생성하였습니다. 이 마스크는 피부 영역의 픽셀만 선택할 수 있도록 해주며, 이를 통해 이미지에서 피부의 RGB 값을 추출했습니다. 추출된 RGB 값은 노이즈를 제거하고 데이터의 왜곡을 최소화하기 위해 1사분위(Q1)와 3사분위(Q3) 사이의 값들만 사용하여 필터링된 평균값을 도출하였습니다.

2)

<https://aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=83>

6.1.3 색상 공간 변환 및 퍼스널 컬러 분류

계산된 피부의 평균 RGB 값은 HSV 및 Lab 색상 공간으로 변환하였습니다. 특히, HSV 색상 공간에서 V(Value)와 S(Saturation), Lab 색상 공간에서 b 채널 값을 추출하였습니다. 이 값을 기반으로 K-means 클러스터링을 활용하여 사용자의 피부색에 따른 퍼스널 컬러 유형을 예측할 수 있는 모델을 구축하였습니다.

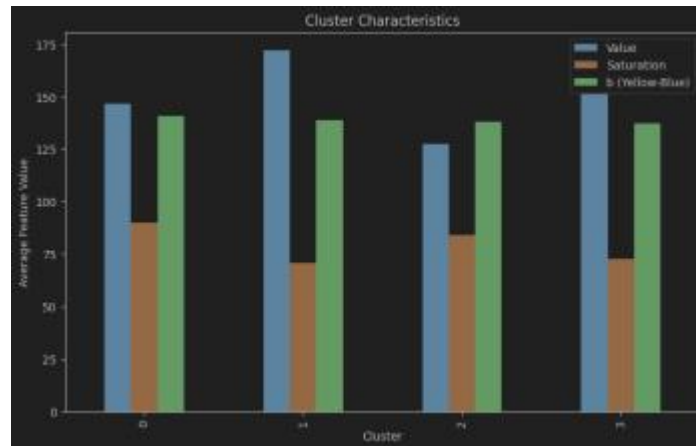


그림 4. 계절별 클러스터

퍼스널 컬러를 예측하기 위한 K-means 클러스터링 모델은 이소영(2019)의 논문 "퍼스널컬러의 톤 유형과 피부색에 따른 조화색의 범주화³⁾"를 참고하여 설계되었습니다. 해당 논문은 퍼스널 컬러를 봄, 여름, 가을, 겨울로 나누고, 각 계절에 따른 피부색의 RGB 값을 도출하여 퍼스널 컬러를 분류하는 기준을 제시합니다.

< 표 7 > 퍼스널컬러 유형 별 피부색

연구자 (연도)	피부색							
	봄		여름		가을		겨울	
김유열 외 (2018)	고명도, 고채도	고명도, 고채도	고명도, 저채도	저명도, 저채도	저명도, 저채도	저명도, 저채도	저명도, 저채도	저명도, 저채도
김연경 (2018)	이미지		이미지		이미지		이미지	
	RGB		RGB		RGB		RGB	
	253, 211, 168	255, 202, 149	253, 231, 174	255, 239, 192	253, 221, 150	247, 208, 152	255, 220, 147	242, 206, 148
	이미지		이미지		이미지		이미지	
	RGB		RGB		RGB		RGB	
	253, 197, 161	252, 204, 130	254, 217, 170	254, 230, 122	249, 201, 128	212, 169, 101	247, 207, 121	216, 173, 102

<다음페이지에 계속>

그림 5. 퍼스널컬러 유형 별 피부색

이 논문에 제시된 각 계절별 대표 피부색 RGB 값과 색상 특징은 모델의 학습에 활용되었습니다. 예를 들어, 봄 타입은 고명도와 고채도의 따뜻한 색조를 지닌 피부색을 의미하고, 겨울 타입은 저명도와 고채도의 차가운 색조를 가진 피부색을 의미합니다.

3) 이소영. "퍼스널컬러의 톤 유형과 피부색에 따른 조화색의 범주화." 국내석사학위논문 홍익대학교 문화정보정책대학원, 2019. 30.

6.2 얼굴형 측정

6.2.1 데이터 전처리

사용된 데이터셋은 Kaggle의 Face Shape Dataset⁴⁾으로, 총 5000개의 이미지로 구성됩니다. 각 얼굴형(하트형, 긴형, 타원형, 둥근형, 사각형)별로 800장의 훈련용 이미지와 200장의 테스트용 이미지가 포함되어 있습니다. 해당 이미지들은 opencv에서 제공하는 CascadeClassifier를 사용하여 얼굴부분만을 추출하였고, 이미지 크기를 380x380 크기로 조정하였습니다. 그 이유는 6.2.2에서 사용된 EfficientNetB4가 380x380의 input shape를 가지고 있기 때문입니다⁵⁾. 5000장의 이미지는 훈련하기에 부족하다고 판단되어 데이터를 증강하였습니다. 데이터 증강은 원본 사진을 수평으로 뒤집은 사진과 (-20, 20)도 범위에서 회전시킨 이미지들을 추가하였습니다.

또한, 사진의 픽셀값 범위는 0~255 범위에 있으므로, 전체 이미지를 R, G, B 평균값과 표준편차를 기반으로 아래와 같이 정규화하였습니다.

$$p_i = \frac{p_i - mean_i}{std_i}, i \in r, g, b$$

여기서 $p_i, mean_i, std_i$ 는 각각 i 색상 픽셀 중 하나, 전체 사진의 i 색상 값 전체 평균, 전체 사진의 i 색상 값 전체 표준편차입니다.

6.2.2 모델 구성

모델 구성은 imagenet 데이터로 사전학습된 EfficientNetB4⁶⁾ 네트워크를 기반으로 구성되었습니다. efficientnetb4에서 globalaveragepooling2d를 적용하고 출력되도록 설정하였으며, efficientnetb4에서 나온 출력값을 Dense층을 통해 5개의 얼굴형으로 분류하였습니다.

4) <https://www.kaggle.com/datasets/niten19/face-shape-dataset>

5) https://keras.io/examples/vision/image_classification_efficientnet_fine_tuning/

6) ingxing Tan and Quoc V. Le. 'EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks'. 2019 <https://arxiv.org/abs/1905.11946>

6.2.3 모델 학습

6.2.3.1 EfficientNetB4를 제외한 나머지만 학습

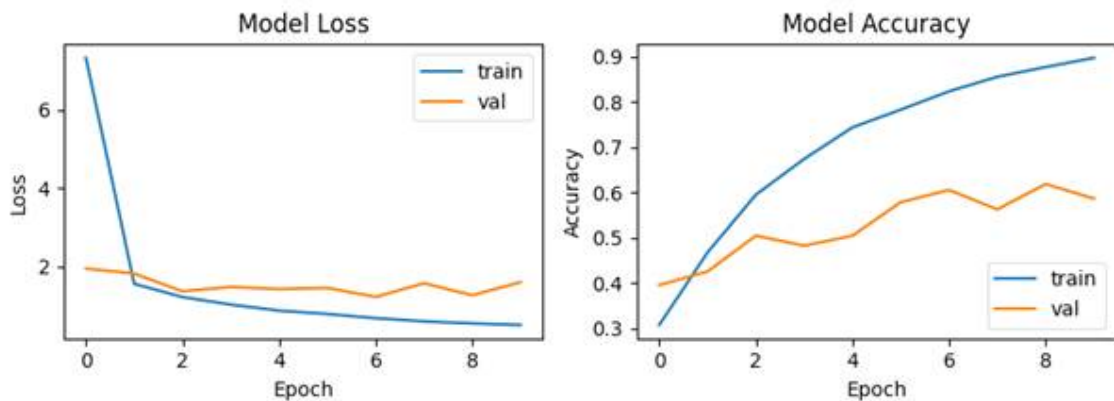


그림 6. EfficientNetB4를 제외한 나머지 학습 그래프

사전학습된 EfficientNetB4의 가중치는 수정하지 않고, 나머지 부분만 학습하도록 하였습니다. 10 epoch 학습시켰으며, adam optimizer(learning rate = 0.001)를 사용하였습니다. 마지막 에포크에서 훈련 정확도는 0.9737, 테스트 정확도는 0.6520이 나왔습니다.

6.2.3.2 전체 학습

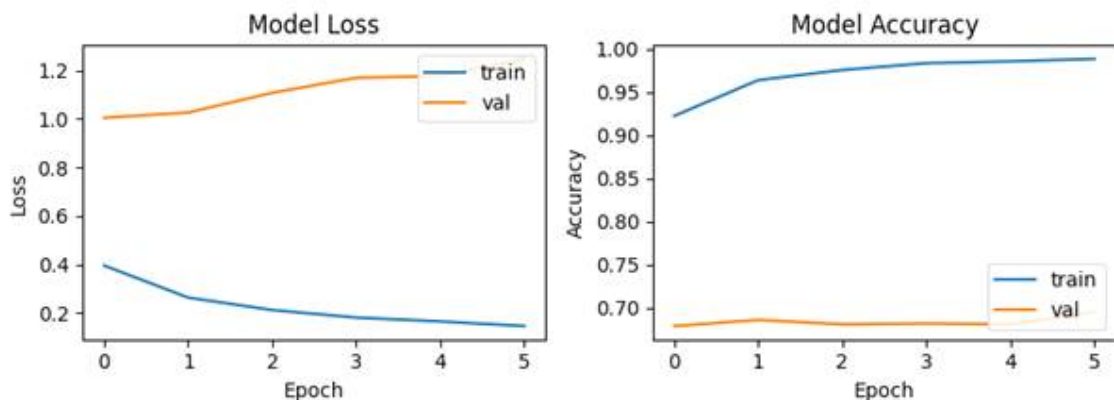


그림 7. EfficientNetB4를 포함한 학습 그래프

EfficientNetB4를 제외하고 학습한 후에, EfficientNetb4를 포함하여 학습하였습니다. 기존 학습된 가중치의 과도한 수정을 방지하기 위해 학습률은 0.0001로 설정하였습니다. 마지막 에포크에서, 훈련 정확도는 0.9978, 테스트 정확도는 0.7452가 나왔습니

다. earlystopping을 적용하였고, restore_best_weights=True로 설정하였기 때문에 훈련 정확도 0.9994, 테스트 정확도 0.7543 모델이 저장되었습니다.

6.3 체형 측정

HumanBodyShape Classification Based on a Single Image⁷⁾를 참고하여 제작하였습니다.

6.3.1 Keypoint 추출

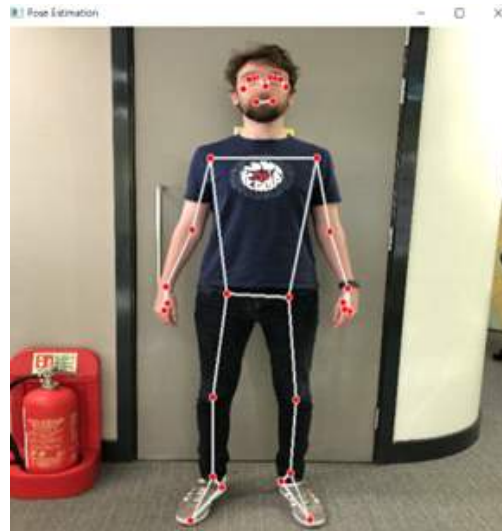


그림 8. MediaPipe의 pose landmark에서 추출하는 신체 부위

체형 측정은 mediapipe 라이브러리의 pose landmark 기능을 사용하여 신체 부위의 keypoint 좌표를 추출하는 방식으로 이루어졌습니다. 이 과정에서 어깨, 엉덩이, 무릎 등의 좌표를 추출하고, 각 부위별로 비율을 계산하여 허리의 위치를 예측하였습니다. 한국인의 신체 데이터를 활용하여 각 신체치수 대비 허리높이 비율을 계산하고, 이를 기반으로 허리 위치를 추정하였습니다.

6.3.2 허리 위치 추측

한국인의 신체 데이터를 정리한 사이즈코리아의 데이터를 사용하였습니다. 여기에는 한국인의 연령별 및 성별 어깨높이, 엉덩이높이, 허리높이, 무릎높이 등이 정리되어 있습니다. 체형별로 정리되어있어 각 체형의 평균 어깨높이, 엉덩이높이, 허리높이, 무릎높이를 구하고, 어깨높이 대비 허리높이, 엉덩이높이 대비 허리높이, 무릎높이 대비 허리높이, (어깨높이+엉덩이높이) 대비 허리높이, (어깨높이+엉덩이높이+무릎높이) 대비 허리높이를 구하여 어느 것이 허리높이를 구하는데 적합한지 알아보았습니다.

7) Cameron Trotter, Filipa Peleja, Dario Dotti, Alberto de Santos. Human Body Shape Classification Based on a Single Image. 2023. <https://arxiv.org/abs/2305.18480>

단위: mm, 평균 (표준편차)	20대 남성	30대 남성	40대 남성	50대 남성	60대 남성
평균 어깨높이	1398.67	1385.75	1364	1348.75	1339
평균 엉덩이높이	849.33	834.5	819.25	805	804
평균 허리높이	1070.33	1054.24	1033	1024.25	1012.75
평균 무릎높이	443	438.25	431.5	422.25	424.5
허리높이/어깨높이	0.765 (0.000042)	0.761 (0.00305)	0.757 (0.00648)	0.759 (0.01437)	0.756 (0.01389)
허리높이/엉덩이높이	1.26 (0.00393)	1.263 (0.0096)	1.261 (0.01699)	1.273 (0.04887)	1.261 (0.04218)
허리높이/무릎높이	2.416 (0.01572)	2.406 (0.02015)	2.395 (0.0525)	2.427 (0.08395)	2.388 (0.09247)
허리높이/어깨+ 엉덩이)	0.476 (0.00072)	0.475 (0.0023)	0.473 (0.00475)	0.476 (0.01182)	0.473 (0.01073)
허리높이/(어깨+ 엉덩이+무릎)	0.398 (0.00091)	0.397 (0.00202)	0.395 (0.00471)	0.398 (0.01047)	0.395 (0.00995)

표 1. 남성의 연령별 신체치수와 각 신체치수 대비 허리높이

단위: mm, 평균 (표준편차)	20대 여성	30대 여성	40대 여성	50대 여성	60대 여성
평균 어깨높이	1289.5	1268.25	1260	1249.25	1223
평균 엉덩이높이	778.75	756.75	748.25	744.56	729.5
평균 허리높이	995	971.75	963.25	951.31	927.5
평균 무릎높이	407.75	398	397.25	393.81	386.5
허리높이/어깨높이	0.772 (0.00317)	0.766 (0.00555)	0.764 (0.00509)	0.76 (0.00733)	0.758 (0.00971)
허리높이/엉덩이높이	1.278 (0.00395)	1.284 (0.00729)	1.287 (0.00729)	1.274 (0.01698)	1.271 (0.01685)
허리높이/무릎높이	2.44 (0.00729)	2.442 (0.0117)	2.425 (0.00835)	2.413 (0.00431)	2.4 (0.02996)
허리높이/어깨+ 엉덩이)	0.481 (0.00092)	0.48 (0.00219)	0.48 (0.00236)	0.476 (0.00504)	0.475 (0.00583)
허리높이/(어깨+ 엉덩이+무릎)	0.402 (0.00071)	0.401 (0.0018)	0.4 (0.00151)	0.398 (0.00343)	0.397 (0.00482)

표 2. 여성의 연령별 신체치수와 각 신체치수 대비 허리높이

표준편차가 상대적으로 가장 낮은 (어깨높이 + 엉덩이높이) 대비 허리높이 평균값을 사용하여 허리 위치를 예측하기로 하였습니다.

허리높이를 구하는 과정은 다음과 같습니다.

1. 어깨와 엉덩이, 그리고 발가락의 keypoint의 좌표를 구한다.
2. (발가락 keypoint의 y좌표 - 어깨 keypoint의 y좌표)를 구하여 어깨높이를 구한다.
3. (발가락 keypoint의 y좌표 - 엉덩이 keypoint의 y좌표)를 구하여 엉덩이높이를 구한다.
4. 표1과 표2에서 구한 비율과 (어깨높이+엉덩이높이)를 곱하여 허리의 y좌표를 구

한다.

6.3.3 가슴 위치 추측

가슴 위치는 한국인의 평균 가슴높이에 대한 데이터가 충분하지 않기 때문에, 어깨 좌표에서 어깨높이의 약 10% 정도의 길이를 아래로 이동하여 가슴의 위치를 추정하였습니다.

6.3.4 배경 제거 및 길이 측정

기존에는 canny edge detection을 사용하여 신체와 배경의 경계부분을 찾으려고 하였으나, 착용하고 있는 옷의 무늬 등으로 인해 측정이 어려워 rembg라는 배경 제거 라이브러리를 사용하여 배경을 제거하였습니다.

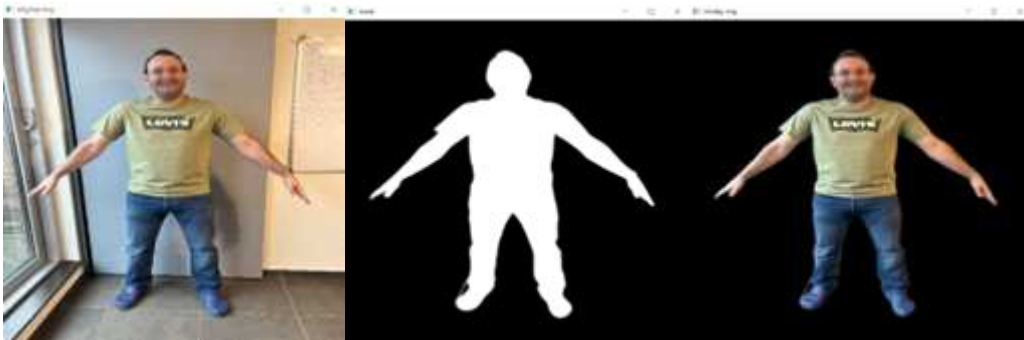


그림 9. 원본 사진(좌), 물체 부분을 하얀색으로 감지(중), 배경 제거된 사진(우)

배경을 제거한 후, 측정된 신체 부위의 좌표를 사용하여 길이를 측정하였습니다. 팔을 붙인 상태에서는 팔의 너비까지 허리와 가슴너비로 측정되기 때문에 그림 10과 같이 팔을 벌린 사진으로 측정하였습니다.

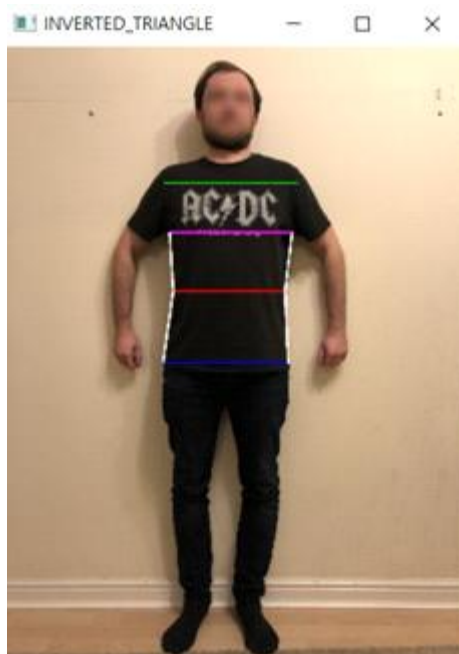


그림 10. 어깨, 가슴, 허리, 엉덩이 너비 측정

6.3.5 체형 측정

측정된 어깨 너비, 가슴너비, 허리너비, 엉덩이너비에 따라 다음과 같이 분류하였습니다.

기준	분류
어깨너비>엉덩이너비 and 엉덩이너비>허리너비 and 가슴너비>엉덩이너비	사다리꼴형
허리너비>어깨너비 and 허리너비>엉덩이너비	둥근형
어깨너비>엉덩이너비 and 어깨너비>가슴너비	역삼각형
엉덩이너비>어깨너비 and 엉덩이너비>가슴너비	삼각형
그외	사각형

표 3. 남성의 체형 분류 기준

기준	분류
어깨너비>허리너비 and 엉덩이너비>허리너비	모래시계형
허리너비>어깨너비 and 허리너비>엉덩이너비	둥근형
어깨너비>엉덩이너비 and 어깨너비>가슴너비	역삼각형
엉덩이너비>어깨너비 and 엉덩이너비>가슴너비	삼각형
그외	사각형

표 4. 여성의 체형 분류 기준

6.4 패션 추천

6.4.1 RandomForest를 활용한 추천시스템

6.4.1.1 데이터 전처리

데이터는 ai-hub에 연도별 패션 선호도 파악 및 추천 데이터⁸⁾의 2019년 데이터를 사용하였습니다. 해당 데이터는 옷 착용 이미지(jpg)와 해당 이미지를 응답자가 평가한 자료(json 형태)로 구성되어있습니다. 평가 자료는 응답자 아이디, 평가한 이미지 이름, 패션 스타일, 이미지 성별, 평가내역(선호여부(1~4점), 어울리는 계절, 적합한 착용상황, 핏, 색깔(어두움, 밝음), 도시적 여부, 세련됨 여부 등 24가지 질문에 대한 응답), 응답자 정보(응답자 성별, 연령, 결혼여부, 선호하는 패션스타일 등)으로 구성되어있다. 각 사용자마다 선호하는 패션스타일은 아래 표1과 같이 패션스타일1, 패션스타일2, 패션스타일3, 패션스타일4, 패션스타일5로 정리되어있으며, 각 패션스타일마다 하나씩 골라 총 5개의 선호하는 패션스타일을 고르게 된다.

패션스타일	패션스타일1	패션스타일2	패션스타일3	패션스타일4	패션스타일5
1	화려하고 독특한	남성적/여성적인	전통적인	포멀한	활발한
2	무난하고 평범한	중성적인	트렌디한	캐주얼한	점잖은

8) <https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=71446>

비고	1, 2중 택1	1, 2중 택1	1, 2중 택1	1, 2중 택1	1, 2중 택1
----	----------	----------	----------	----------	----------

표 5. 응답자의 선호하는 패션스타일 선택

데이터 전처리 과정에서는 json 데이터를 csv로 변환하여 사용자 정보, 옷 정보, 평가 정보를 분리하였다. 옷에 대해 여러 사용자가 평가한 경우, 최빈값을 채택하는 방식으로 해당 옷의 최종 평가를 결정하였습니다. 이후 사용자가 선호하는 패션 스타일에 따라 퍼스널 컬러, 얼굴형, 체형을 유추하였습니다. 아래 표를 사용하여 스타일별로 퍼스널 컬러와 신체적 특성을 부여하고, 가장 빈도가 높은 특성을 최종 선택하였습니다.

선호하는 패션스타일	퍼스널컬러	얼굴형	체형
화려하고 독특한	봄, 겨울	타원형, 하트형	역삼각형, 사다리꼴형
무난하고 평범한	여름, 가을	둥근형, 사각형	둥근형, 직사각형
남성적인	가을, 겨울	긴형, 사각형	역삼각형, 사다리꼴형
중성적인	봄, 여름	타원형	직사각형
전통적인	가을, 겨울	둥근형, 사각형, 긴형	사다리꼴형, 둥근형
트렌디한	봄, 여름	하트형, 타원형	역삼각형, 삼각형
포멀한	가을, 겨울	긴형, 사각형	직사각형, 사다리꼴형
캐주얼한	봄, 여름	둥근형, 하트형	둥근형, 삼각형
활발한	봄, 여름	둥근형, 하트형	삼각형, 역삼각형
점잖은	가을, 겨울	사각형, 타원형, 긴형	사다리꼴형, 직사각형

표 6. 남성 응답자에 퍼스널 컬러, 얼굴형, 체형 부여

선호하는 패션스타일	퍼스널컬러	얼굴형	체형
화려하고 독특한	봄, 겨울	타원형, 하트형	역삼각형, 모래시계형
무난하고 평범한	여름, 가을	둥근형, 사각형	둥근형, 직사각형
남성적인	가을, 겨울	긴형, 사각형	모래시계형
중성적인	봄, 여름	타원형	직사각형
전통적인	가을, 겨울	둥근형, 사각형, 긴형	직사각형, 둥근형
트렌디한	봄, 여름	하트형, 타원형	역삼각형, 삼각형
포멀한	가을, 겨울	긴형, 사각형	직사각형, 모래시계형
캐주얼한	봄, 여름	둥근형, 하트형	둥근형, 삼각형
활발한	봄, 여름	둥근형, 하트형	삼각형, 역삼각형
점잖은	가을, 겨울	사각형, 타원형, 긴형	모래시계형, 직사각형

표 7. 여성 응답자에 퍼스널 컬러, 얼굴형, 체형 부여

6.4.1.2 Random Forest를 활용한 학습

학습에 사용된 데이터는 사용자 특성(퍼스널컬러, 얼굴형, 체형)과 옷의 특성을 feature로, 평점을 target으로 구성하였습니다. 데이터는 OneHot Encoding을 통해 전처리되었고, scikit-learn의 RandomForestRegressor를 사용하여 학습하였습니다. 학습 결과의 RMSE는 남성 데이터의 경우 훈련 데이터에서 0.304, 테스트 데이터에서 0.759로, 여성 데이터의 경우 훈련 데이터에서 0.312, 테스트 데이터에서 0.773으로 나타났습니다.

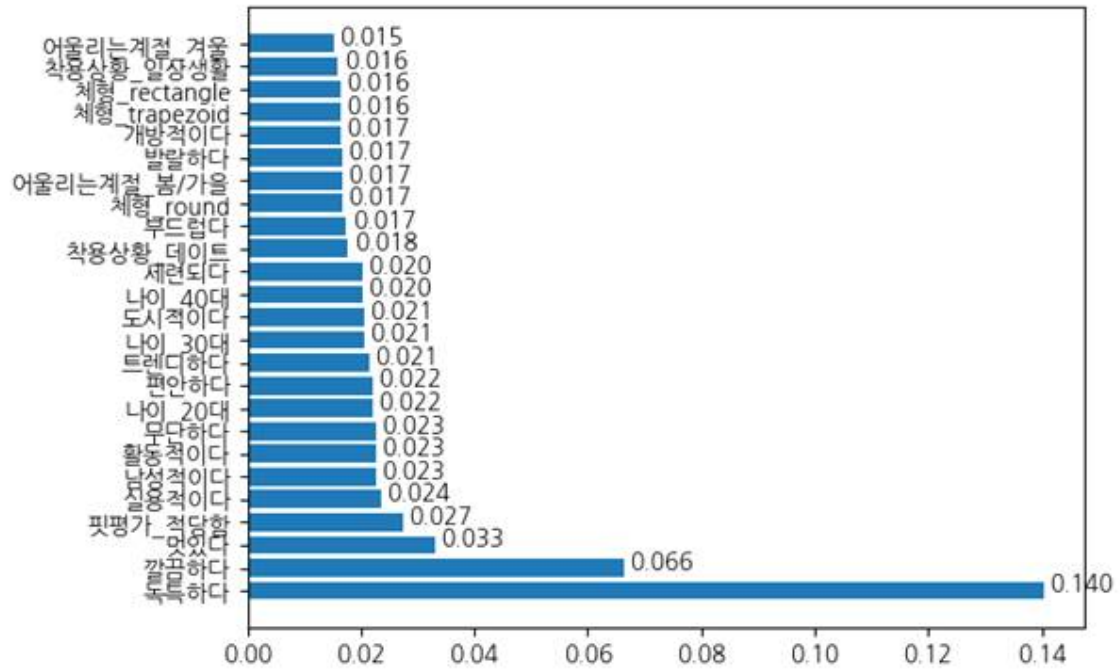


그림 11. 남성 옷 추천에 있어 모델이 가장 중요하게 생각하는 feature

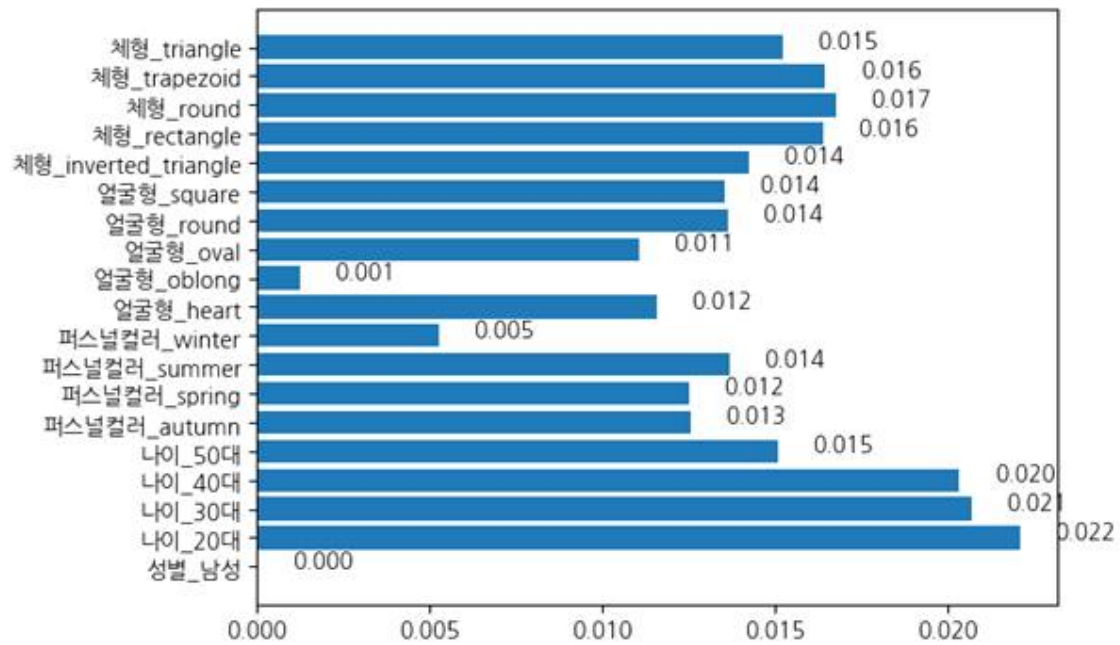


그림 12. 남성 옷 추천에 있어 모델이 생각하는 사용자 특성 중요도

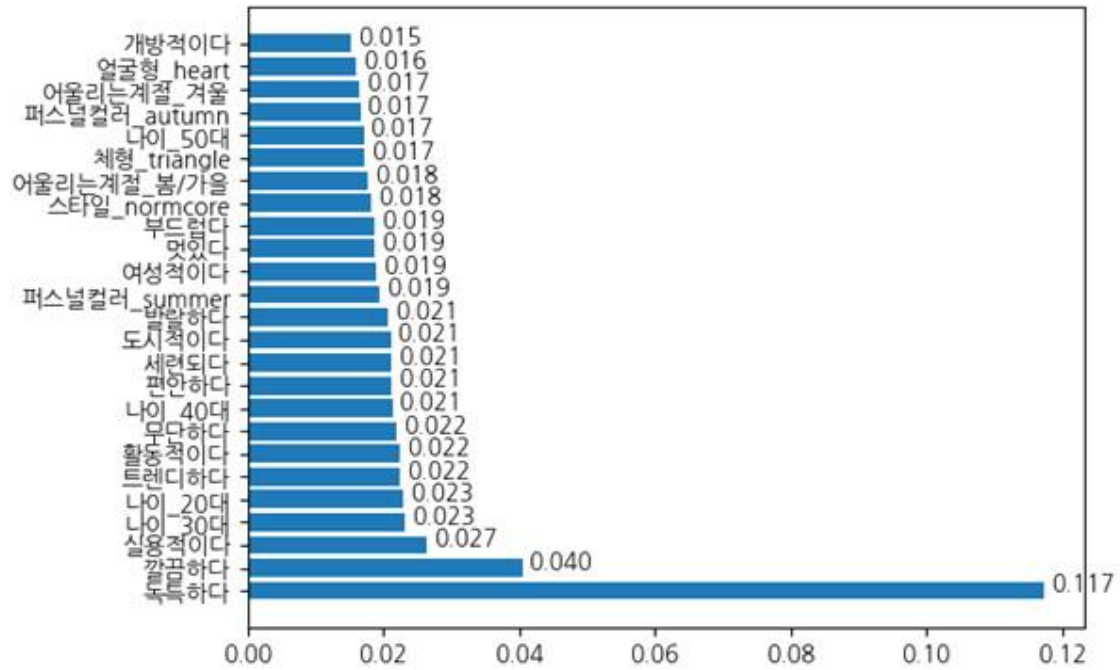


그림 13. 여성 옷 추천에 있어 모델이 가장 중요하게 생각하는 feature

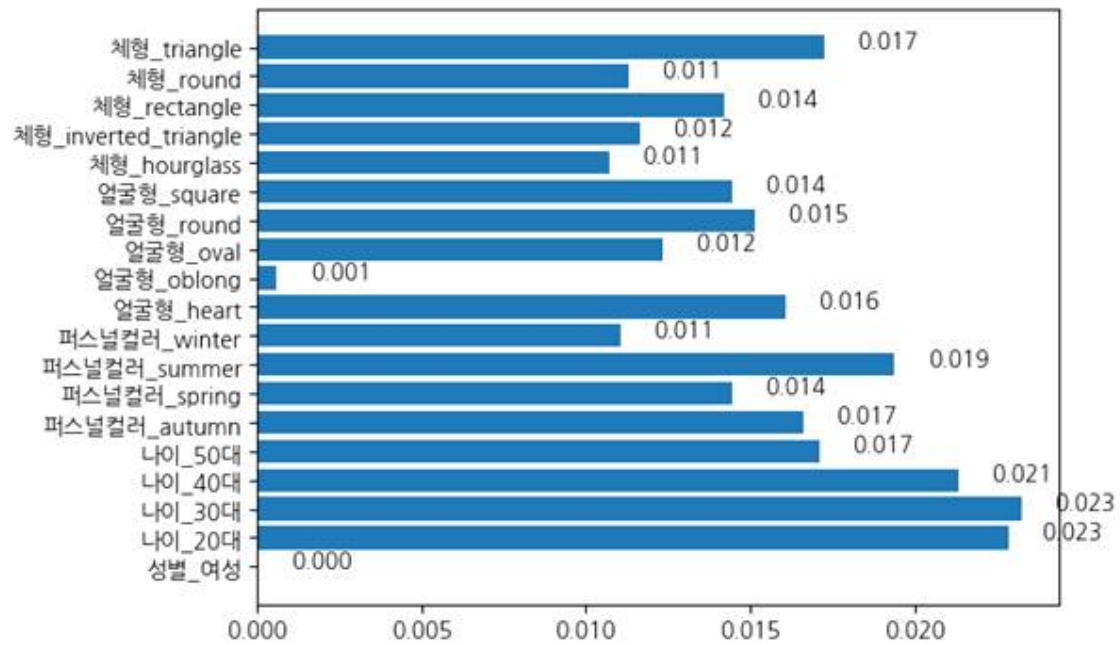


그림 14. 여성 옷 추천에 있어 모델이 생각하는 사용자 특성 중요도

성별/데이터	훈련 데이터	테스트 데이터
남	0.304	0.759
여	0.312	0.773

표 8. RandomForest를 사용한 훈련 결과

만들어진 모델을 사용하여 사용자 특성 - 옷 특성에 대해 평점 예측을 하여 평점이 높은 순서대로 추천을 해보면 다음과 같이 표9, 표10의 결과가 나온다.

옷 이름	예측 평점(4점 만점)
W_81578_19_normcore_M.jpg	3.780000
W_63901_19_normcore_M.jpg	3.770000
W_54264_19_normcore_M.jpg	3.760000
T_17451_19_normcore_M.jpg	3.760000
W_81468_19_normcore_M.jpg	3.750000
W_01418_19_normcore_M.jpg	3.743333
T_19532_19_normcore_M.jpg	3.740000
T_17134_19_normcore_M.jpg	3.700000
W_31826_19_normcore_M.jpg	3.700000
T_17087_19_normcore_M.jpg	3.683333

표 9. 20대 남성, 봄 퍼스널컬러, 둥근 얼굴형, 둥근 체형에 대한 평점 예측 결과

옷 이름	예측 평점(4점 만점)
W_07708_19_normcore_W.jpg	3.710000
W_05429_19_normcore_W.jpg	3.710000
T_06247_19_normcore_W.jpg	3.706667
T_06252_19_normcore_W.jpg	3.675000
W_69663_19_normcore_W.jpg	3.656667
T_14752_19_normcore_W.jpg	3.626667
T_15674_19_normcore_W.jpg	3.625000
T_05226_19_normcore_W.jpg	3.610000
W_37378_19_normcore_W.jpg	3.607500
W_04139_19_normcore_W.jpg	3.606667

표 10. 20대 여성, 봄 퍼스널컬러, 타원 얼굴형, 직사각형 체형에 대한 평점 예측 결과

6.4.1.3 추천 품질에 대한 피드백

자문의견서에 기재된 추천 품질에 대한 피드백을 제공하는 것을 구현하였습니다. 구현방법은 6.4.1.2에 있는 randomforest를 사용한 모델을 사용하여 다음과 같이 구성 하였습니다.

1. 추천 시 모델이 예측한 평점과 실제 사용자가 평가한 평점을 함께 반영하여 최종 평점을 산출한다. 모델의 예측 평점을 70%, 실제 사용자의 평점을 30% 반영하는 방식이다.
2. 사용자가 별점을 매기면 해당 데이터를 기존 훈련 데이터에 추가하고, 새로운 RandomForest 모델을 학습시킨다. 사용자가 매긴 평점은 사용자별로 기록되고, 실제 사용자의 평점에 반영된다.

표9와 표10의 추천결과를 6.4.1.3-1을 반영하여 추천하면 각각 표11과 표12가 된다.

옷 이름	예상평점	사용자 평점	종합평점
W_81578_19_normcore_M.jpg	3.78	4	3.846
W_63901_19_normcore_M.jpg	3.77	4	3.839
T_17451_19_normcore_M.jpg	3.76	4	3.832
W_54264_19_normcore_M.jpg	3.76	4	3.832
W_81468_19_normcore_M.jpg	3.75	4	3.825
T_19532_19_normcore_M.jpg	3.74	4	3.818
W_31826_19_normcore_M.jpg	3.7	4	3.79
T_17134_19_normcore_M.jpg	3.7	4	3.79
T_17087_19_normcore_M.jpg	3.6833	4	3.778
T_19542_19_normcore_M.jpg	3.68	4	3.776

표 11. 20대 남성, 봄 퍼스널컬러, 둥근 얼굴형, 둥근 체형에서 평점 예측

옷 이름	예상평점	사용자 평점	종합평점
W_69663_19_normcore_W.jpg	3.6567	4	3.7597
T_14752_19_normcore_W.jpg	3.6267	4	3.7387
T_15674_19_normcore_W.jpg	3.625	4	3.7375
T_05226_19_normcore_W.jpg	3.61	4	3.727
W_37378_19_normcore_W.jpg	3.6075	4	3.72523
W_04139_19_normcore_W.jpg	3.6067	4	3.7247
T_13333_19_normcore_W.jpg	3.6	4	3.72
T_05105_19_normcore_W.jpg	3.6	4	3.72
T_06242_19_lounge_W.jpg	3.6	4	3.72
T_05065_19_normcore_W.jpg	3.59	4	3.713

표 12. 20대 20대 여성, 봄 퍼스널컬러, 둥근 얼굴형, 둥근 체형에서 평점 예측

이제 표11과 표12에서 추천시스템이 가장 추천하는 옷에 대하여 평점을 1점으로 평가한 후, 다시 추천하면 표13 및 표14와 같습니다.

옷 이름	예상평점	사용자 평점	종합평점
W_54264_19_normcore_M.jpg	3.77	4	3.839
T_17451_19_normcore_M.jpg	3.77	4	3.839
W_63901_19_normcore_M.jpg	3.76	4	3.832
W_81468_19_normcore_M.jpg	3.76	4	3.832
W_31826_19_normcore_M.jpg	3.705	4	3.7935
T_17134_19_normcore_M.jpg	3.7	4	3.79
W_17682_19_normcore_M.jpg	3.68	4	3.776
W_52075_19_normcore_M.jpg	3.67	4	3.769
W_25216_19_normcore_M.jpg	3.67	4	3.769
W_92952_19_normcore_M.jpg	3.67	4	3.769

표 13. 20대 남성, 봄 퍼스널컬러, 둥근 얼굴형, 둥근 체형에서 피드백 후 추천 결과

옷 이름	예상평점	사용자 평점	종합평점
T_05065_19_normcore_W.jpg	3.66	4	3.7612
W_37378_19_normcore_W.jpg	3.6525	4	3.7567
T_06242_19_lounge_W.jpg	3.63	4	3.741
T_13333_19_normcore_W.jpg	3.63	4	3.741
T_14752_19_normcore_W.jpg	3.6067	4	3.7247
T_05226_19_normcore_W.jpg	3.6	4	3.72
W_04139_19_normcore_W.jpg	3.5967	4	3.7177
W_90330_19_normcore_W.jpg	3.59	4	3.713
T_15674_19_normcore_W.jpg	3.58	4	3.7095
T_05123_19_lounge_W.jpg	3.58	4	3.706

표 14. 20대 여성, 봄 퍼스널컬러, 둥근 얼굴형, 둥근 체형에서 피드백 후 추천 결과

6.4.2 콘텐츠 기반 협업필터링 추천시스템

6.4.2.1 데이터 전처리

데이터는 AI-hub의 K-Fashion 이미지⁹⁾를 사용하였습니다. 해당 데이터는 120만 장의 옷 착용 사진과 해당 옷의 특징을 json 파일로 정리되어있습니다. 이를 파싱하여 상의, 하의, 아우터, 원피스 등의 카테고리로 구분하고, csv 파일로 정리하였습니다.

옷아이디	색상	소매기장	서브색상	스타일	서브스타일	카테고리	옷깃	디테일	소재	프린트	핏	기장	넥라인
885370	화이트	민소매	레드	협합	스트리트	탑			우븐	스트라이프	타이트	크롭	
918045	화이트	캡		협합	키치	티셔츠			우븐	무지	루즈	노멀	라운드넥
918145	화이트	반팔	블랙	협합	키치	티셔츠			우븐	무지 레터링	오버사이즈	몸	라운드넥
994937	화이트	7부소매	레드	협합		티셔츠			우븐	레터링 무지	루즈	노멀	라운드넥
994957	화이트	7부소매	레드	협합		티셔츠			우븐	레터링 무지	루즈	노멀	라운드넥

그림 15. 정리된 상의 데이터의 일부

9)

https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=real_m&dataSetSn=51

6.4.2.2 Tf-Idf Vectorizing

Tf-Idf Vectorizing은 키워드의 중요도를 나타내는 기법으로, 옷의 각 특성을 하나의 문자열로 합친 후 scikit-learn의 Tf-Idf Vectorizer를 사용하여 벡터화하였습니다. 이 과정에서 l2 정규화를 적용하여 과적합을 방지하였습니다.

```
top['combine_features'].tail()
✓ 0.0s
```

77004	화이트 민소매 레드 힙합 스트리트 탑 우븐 스트라이프 타이트 크롭
77005	화이트 캡 힙합 키치 티셔츠 우븐 무지 루즈 노멀 라운드넥
77006	화이트 반팔 블랙 힙합 키치 티셔츠 우븐 무지 레터링 오버사이즈 롱 라운드넥
77007	화이트 7부소매 레드 힙합 티셔츠 우븐 레터링 무지 루즈 노멀 라운드넥
77008	화이트 7부소매 레드 힙합 티셔츠 우븐 레터링 무지 루즈 노멀 라운드넥

Name: combine_features, dtype: object

그림 16. 옷의 특성을 하나의 문자열로 합침

```
tfidf_vectorizer = TfidfVectorizer(analyzer='word', norm='l2')
tfidf_matrix = tfidf_vectorizer.fit_transform(top['combine_features'])
tfidf_vectorizer.vocabulary_.items()
✓ 0.7s
```

dict_items([('베이지', 54), ('긴팔', 10), ('라벤더', 27), ('화이트', 160), ('그레이', 7),

그림 17. Tf-Idf Vectorizer로 벡터화 하는 코드

```
clothes_feature_from_user = ['카키 긴팔 스트리트 스포티 티셔츠 저지 그래픽 오버사이즈 롱 라운드넥']
tfidf_new = tfidf_vectorizer.transform(clothes_feature_from_user)
tfidf_new.shape
✓ 0.0s
```

(1, 165)

```
print(tfidf_new)
✓ 0.0s
```

<Compressed Sparse Row sparse matrix of dtype 'float64'
with 9 stored elements and shape (1, 165)>

Coords	Values
(0, 6)	0.411807268796893
(0, 10)	0.18103041023176147
(0, 28)	0.19383617625670918
(0, 80)	0.15070333416540335
(0, 85)	0.468280597953695
(0, 98)	0.41315414934882994
(0, 107)	0.24166890012944187
(0, 116)	0.49474747884761094
(0, 132)	0.2102303763091945

그림 18. 테스트로, '카키 긴팔 스트리트 스포티 티셔츠 저지 그래픽 오버사이즈 롱 라운드넥' 을 벡터화한 결과

6.4.2.3 코사인 유사도 구하기

코사인 유사도는 내적공간의 두 벡터간 각도의 코사인 값을 이용하여 측정된 유사도입니다. $X=(x_1,...,x_n)$, $Y=(y_1,...,y_n)$ 을 두 문서의 벡터쌍이라고 하면 코사인 유사도는 아래와 같이 구할 수 있습니다.

$$Cosine(X,Y)=\frac{\sum x_iy_i}{\sqrt{\sum x_i^2}\sqrt{\sum y_i^2}}$$

scikit-learn의 cosine_similarity 함수를 사용하여 벡터화된 옷 특성 간의 유사도를 계산하고, 유사도가 높은 순서대로 추천하였습니다.

옷번호	스타일	옷특징	유사도
1038491	스트리트	카키 긴팔 스트리트 스포티 티셔츠 저지 그래픽 오버사이즈 롱 라운드넥	1
37620	스트리트	카키 긴팔 화이트 스트리트 티셔츠 저지 그래픽 오버사이즈 롱 라운드넥	0.8614
1036139	스트리트	카키 긴팔 스트리트 스포티 티셔츠 우븐 그래픽 루즈 롱 라운드넥	0.8309
190270	스트리트	스트리트 스포티 티셔츠 저지 그래픽 오버사이즈 롱	0.8276
1213067	스트리트	카키 긴팔 스트리트 스포티 티셔츠 드롭숄더 저지 무지 오버사이즈	0.8208
735440	스트리트	블랙 긴팔 화이트 스트리트 스포티 티셔츠 저지 그래픽 오버사이즈 롱 라운드넥	0.8035
735430	스트리트	블랙 긴팔 화이트 스트리트 스포티 티셔츠 저지 그래픽 오버사이즈 롱 라운드넥	0.8035
735410	스트리트	블랙 긴팔 화이트 스트리트 스포티 티셔츠 저지 그래픽 오버사이즈 롱 라운드넥	0.8035
37630	스트리트	카키 긴팔 화이트 스트리트 티셔츠 슬릿 저지 그래픽 오버사이즈 롱 라운드넥	0.7719
397965	스트리트	카키 긴팔 블랙 스트리트 티셔츠 드롭숄더 저지 레터링,그래픽 오버사이즈 롱 라운드넥	0.7459

표 15. ‘카키 긴팔 스트리트 스포티 티셔츠 저지 그래픽 오버사이즈 롱 라운드넥’ 특성과 유사한 상의

6.4.2.4 사용자의 신체적 특징에 어울리는 옷 특성 연결

사용자의 퍼스널컬러, 얼굴형, 체형에 어울리는 옷 특성을 연결하였습니다. 예를 들어 사각형 얼굴에는 어울리는 상의 특성에 라운드넥을 추가하였습니다.

참고문헌

2023년 소비 트렌드 시리즈 - 03 개인맞춤형 서비스,

https://www.mezzomedia.co.kr/data/insight_m_file/insight_m_file_1605.pdf,

2024.05.19

신향선. "퍼스널 컬러시스템에 따른 유형의 분포도와 색채진단 변인에 관한 연구." 국내석사학위논문 건국대학교, 2002.

차호연. "국내외 퍼스널 컬러 진단시스템의 비교." 국내박사학위논문 원광대학교 일반대학원, 2011.

이소영. "퍼스널컬러의 톤 유형과 피부색에 따른 조화색의 범주화." 국내석사학위논문 홍익대학교 문화정보정책대학원, 2019.

AIHub, "K-Fashion 이미지",

<https://aihub.or.kr/aihubdata/data/view.do?dataSetSn=51>, 2024.05.19.

AIHub, "한국인 안면 이미지",

<https://aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=83>, 2024.07.26.

가자마 마사히로 외, '추천 시스템 입문', 김모세 역, 한빛미디어

차루 아가르왈. '추천 시스템'. 박희원 외 역. 에이콘출판

Tan, M., & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 2020.

Trotter, C., Peleja, F., Dotti, D., & de Santos, A. Human Body Shape Classification Based on a Single Image. 2023.